Proceedings of CARSIT 2027

International Conference on Business Innovation, Artificial Intelligence & Emerging Digital Technologies

CARSIT InnovX 2027

Editors	Prof. Dr. Karim Mohammed Rezaul
Publisher	CARSIT
Publication date	03 January 2027

Official Conference Proceedings

Publication & Copyright Information

This volume contains eligible accepted and published papers of CARSIT InnovX 2027. Individual papers retain their stated licence and publication metadata.

Table of Contents

1. A Hybrid Framework for Mpox Diagnosis and Severity Assessment via Symptom Classification and Skin Lesion Segmentation

Md Jewel, Karim Mohammed Rezaul, Kazy Noor e Alam Siddiquee, Md Shabiul Islam, Tajnuva Jannat · Compiled page 4

A Hybrid Framework for Mpox Diagnosis and Severity Assessment via Symptom Classification and Skin Lesion Segmentation

Md Jewel
me.jewel.md@gmail.com
Centre for Applied Research in Software & IT (CARSIT), UK

Karim Mohammed Rezaul
rezababu@gmail.com
University of Westminster, UK

Kazy Noor e Alam Siddiquee
knas11@gmail.com
Military Institute of Science & Technology (MIST), Bangladesh

Md Shabiul Islam
shabiul.islam@mmu.edu.my
Multimedia University, Malaysia

Corresponding Author: Prof. Dr. Md Shabiul Islam

Abstract

Mpox has re-emerged as a global health concern, demanding diagnostic solutions that reflect the complex clinical presentation of the disease. Traditional machine learning and deep learning approaches often isolate either symptom-based or image-based analysis, limiting diagnostic comprehensiveness. To address this gap, we propose a novel hybrid, late-fusion AI-driven framework that integrates structured symptom data with skin lesion image analysis for reliable detection and severity assessment. Separate symptom and image datasets were utilized to develop the respective models. For symptom-based classification, an ensemble of six machine learning models including LightGBM, XGBoost, CatBoost, Logistic Regression, Random Forest, and Support Vector Machine is employed. Concurrently, a majority voting ensemble of seven pre-trained deep learning models, including ResNet50, InceptionV3, MobileNetV2, EfficientNet, VGG16, Xception, and DenseNet, is used for image-based diagnosis. Positive image classifications trigger lesion segmentation using a customized U-Net model to quantify lesion burden. A severity scoring algorithm, designed according to World Health Organization (WHO) clinical guidelines, combines symptom severity and lesion metrics to classify cases as mild, moderate, severe, or serious. The proposed framework provides explainable and clinically interpretable outputs within an internal experimental setting, achieving an AUC of 0.71 for symptom classification, up to 97.00% validation accuracy for image classification, and a Dice coefficient of 0.76 for lesion segmentation. These findings demonstrate the technical feasibility of the proposed late-fusion framework; however, external validation on independent patient-level multimodal datasets is required before claims of clinical generalizability or real-world deployment can be made. Designed as a prototype for future real-world adaptation, our system provides a foundational AI-based tool to support future early intervention strategies.

Introduction

Mpox is an emerging zoonotic disease caused by the Mpox virus (MPXV) which is a member of the Orthopoxvirus genus within the Poxviridae family. Although mpox shares clinical similarities with smallpox, it exhibits lower mortality rates but remains a significant global health concern due to its increasing incidence beyond endemic regions in Central and West Africa. Historically, vaccinia-based smallpox vaccination provided cross-protection, but its cessation has led to increased susceptibility to mpox infections [1]. The primary mode of transmission involves zoonotic spillover from African rodents, while human-to-human transmission occurs via respiratory droplets, direct contact, and contaminated surfaces. Recent outbreaks suggest that high viral loads detected in bodily fluids may indicate sexual transmission as a major contributing factor [2,3]. Two genetically distinct clades of MPXV exist: The Central African (Congo Basin) clade, characterized by higher transmissibility and severity, and the West African clade, which exhibits lower virulence and limited human-to-human spread [4,5]. The clinical presentation of mpox typically follows an incubation period of 7–14 days, progressing through a febrile prodrome with lymphadenopathy to a centrifugal rash that evolves through macular, papular, vesicular, and pustular phases before resolution over 2–4 weeks. Symptoms such as fever, sore throat, rectal pain, and swollen lymph nodes, along with characteristic skin lesions, create diagnostic challenges, particularly in resource-limited settings [6-8]. Current diagnostic methods rely on clinical evaluation, polymerase chain reaction (PCR) assays, and serological tests [9,10]. However, these approaches require specialized laboratories, trained personnel, and extended processing times, limiting their applicability in outbreak settings and low-resource regions. Given these constraints, artificial intelligence (AI)-driven diagnostic tools have emerged as promising alternatives for rapid, scalable, and accessible mpox detection.

1.1 Challenges in Mpox Diagnosis and the Need for AI-driven Solutions

Accurate and early detection of mpox remains challenging, particularly in low-resource environments. Clinical evaluation is often non-specific due to symptom overlap with other viral exanthems such as varicella or herpes. While PCR testing is highly sensitive, its reliance on specialized facilities limits accessibility. Serological testing, though useful for retrospective diagnosis, lacks immediate applicability in outbreak control. Given these limitations, AI-driven diagnostic tools leveraging machine learning (ML) and deep learning (DL) techniques have demonstrated potential in dermatological and infectious disease classification, particularly in skin lesion analysis. However, existing models predominantly focus on image-based classification, overlooking the integration of clinical symptomatology, which is crucial for a comprehensive diagnosis.

1.2 Proposed Hybrid, Late-Fusion AI Framework for Mpox Detection and Severity Assessment

To address these diagnostic challenges, we propose a novel hybrid, late-fusion AI-based framework that integrates both symptom-based and image-based analysis for mpox detection and severity assessment. Our approach leverages:

1. **Symptom-Based Classification:** An ensemble of six ML classifiers including LightGBM, XGBoost, CatBoost, Logistic Regression, Random Forest, and Support Vector Machine (SVM) is employed to predict mpox likelihood based on patient-reported symptoms.
2. **Image-Based Classification:** A deep learning ensemble incorporating ResNet50, InceptionV3, MobileNetV2, EfficientNet, VGG16, Xception, and DenseNet is used for skin lesion classification, employing a majority voting strategy to enhance predictive accuracy.
3. **Lesion Severity Quantification:** For individuals classified as mpox-positive through image-based models, a U-Net-based segmentation approach is employed to quantify lesion burden according to World Health Organization (WHO) lesion count guidelines, while symptom-based severity is calculated separately and subsequently normalized for the final comprehensive assessment.

By aggregating multimodal inputs at the decision level, this framework enhances diagnostic reliability and provides a comprehensive assessment of disease severity, facilitating informed clinical decision-making and personalized patient management.

1.3 Our Key Contributions

Our study presents the following key contributions:

i. **A Hybrid Late-Fusion AI Framework:** We propose a decision-level late-fusion framework that integrates structured symptom-based classification with image-based skin lesion analysis for mpox diagnosis and preliminary severity assessment.
ii. **Dual Ensemble Classification Strategy:** We implement ensemble learning in both branches of the framework. For symptom-based prediction, six machine learning classifiers are combined through a hierarchical stacking strategy. For image-based diagnosis, seven pre-trained convolutional neural networks are integrated using majority voting to improve classification robustness.
iii. **Segmentation-Based Lesion Quantification:** We incorporate a customized U-Net segmentation model to localize mpox lesions and extract lesion-burden indicators, including lesion count and segmented lesion area, for downstream severity scoring.
iv. **Interpretable Guideline-Informed Severity Scoring:** We develop a deterministic and interpretable severity scoring mechanism that combines symptom severity and lesion-burden metrics into a normalized final score. This score provides preliminary severity categorization while remaining transparent and suitable for future clinical calibration.
v. **System-Level Workflow Demonstration:** We demonstrate the complete operational workflow of the proposed framework using a test case, including symptom scoring, image classification, segmentation triggering, lesion quantification, score fusion, threshold sensitivity analysis, and final severity categorization.

By combining symptom-based classification, image-based detection, lesion segmentation, and severity scoring, this study presents a prototype late-fusion framework for mpox diagnosis and preliminary severity assessment, particularly for future development in outbreak-prone and

resource-limited settings. The framework is not intended as a clinically validated diagnostic system at this stage and requires external validation using independent patient-level multimodal datasets.

The rest of the paper is structured as follows: Section 2 reviews related work on mpox classification and Section 3 describes our proposed methodology, including dataset preprocessing, model architecture, and evaluation metrics. Section 4 presents experimental test case, followed by Section 5, which outlines key findings, limitations, and future research directions. Finally, Section 6 concludes the paper with a summary of contributions and potential clinical applications of our framework.

2. Related Work

In recent years, significant advancements have been made in the development of machine learning (ML) and deep learning (DL) models for the detection and classification of Mpox. However, the majority of existing studies primarily focus on image-based analysis, often neglecting the integration of clinical symptom data. Incorporating such data could significantly improve diagnostic accuracy and enhance severity assessment. While previous research has examined the clinical presentation, transmission dynamics, and management of Mpox, a notable gap remains in leveraging multimodal, AI-driven approaches for both detection and severity scoring.

Our work seeks to fill this gap by integrating both image and symptom data into a hybrid late-fusion framework for early detection and severity assessment of Mpox. In this section, we review key studies in the field, highlighting their contributions, limitations, and how our approach extends the existing body of knowledge.

2.1 Deep Learning for Image-Based Mpox Detection

In recent years, deep learning has been widely applied to the detection and classification of Mpox skin lesions. Early studies primarily relied on convolutional neural networks (CNNs) and transfer learning, often incorporating Explainable AI (XAI) techniques such as Grad-CAM and LIME to improve model interpretability [11, 12, 14]. While these works laid a solid foundation for image-based detection, they generally focused on classification accuracy alone and paid limited attention to real-world deployment issues, symptom integration, or disease severity assessment. Similarly, several studies explored mobile applications and patch-based methods to enhance accessibility [15, 17, 30], but these too remained strictly unimodal. Other early efforts focused on transfer learning and diagnostic accuracy but lacked multimodal integration [16, 20, 31, 32].

More recent efforts have introduced increasingly sophisticated architectures. Khan and Iqbal (2024), for example, combined Swin Transformers with residual and spatial CNNs to boost lesion detection performance, although their approach was still limited to image data [13]. Throughout 2024 and 2025, the field moved toward hybrid and attention-based models [18, 19, 33, 34]. By 2026, ensemble learning and attention mechanisms had become prominent. Das et al. (2026) proposed a framework that integrates ensemble modeling with attention layers, reporting improved classification results [43]. Likewise, Zhang et al. (2026) developed an attention-

enhanced hybrid model using DenseNet-121 and EfficientNet-B4 with CBAM attention, achieving strong performance on skin lesion images [46].

There has also been a clear push toward practical, deployable solutions. Askale et al. (2026) introduced a lightweight CNN based on MobileNetV2, specifically designed for resource-limited settings and mobile devices [44]. Sharma et al. (2026) further optimized a lightweight attention-based CNN using genetic algorithms, aiming for both efficiency and accuracy [47]. For real-time applications, Alghoraibi et al. (2025–2026) proposed the ITMA'INN system, which leverages Vision Transformers and MobileViT architectures [48].

In parallel, progress has been made in lesion segmentation. Vandana et al. (2026) presented MpoxSegNet, a model that performs both multiclass segmentation and classification using a modified VGGNet-16 backbone [45]. Onyema et al. (2025–2026) improved segmentation quality by incorporating hair artifact removal techniques within an enhanced U-Net framework [49]. Despite these advances, most existing models remain unimodal focusing either on classification or segmentation and few attempt to translate segmentation outputs into clinically meaningful severity scores based on WHO guidelines.

2.2 Clinical and Epidemiological Studies on Mpox

A large body of clinical and epidemiological literature has documented the symptoms, transmission routes, risk factors, and complications of Mpox [21–23, 26, 27]. Comprehensive reviews have further detailed clinical risk factors and diagnostic overviews [24, 25], yet these remain disconnected from AI-driven solutions. Recent large-scale studies from 2024 and 2025 have provided updated insights into the clinical presentation of newer clades and associated complications [50, 51]. Clinical trials evaluating antivirals such as tecovirimat have also continued to inform treatment protocols [52], while updated WHO guidance and public health responses to Clade Ib have refined severity classification approaches [53, 54]. Although these studies significantly enrich clinical understanding, they rarely connect symptomatic data with automated AI-based diagnostic systems.

2.3 Epidemiological and Preventive Studies

Epidemiological research has examined global transmission patterns and prevention strategies [28, 29]. More recently, machine learning approaches have been applied to outbreak forecasting and vaccine impact assessment [55, 56]. There is also growing interest in using Large Language Models (LLMs) for initial symptom triage [57]. However, these predictive and text-based models generally lack visual confirmation from lesion images. Even recent hybrid AI frameworks developed for broader dermatological conditions [58] have not yet fully integrated symptom classification, image-based detection, lesion segmentation, and severity scoring into a unified system for Mpox.

2.4 Contributions of Our Work

While the studies reviewed above represent meaningful progress in Mpox detection, most remain limited to single-modality (primarily image-based) approaches. Our work addresses this gap by proposing a hybrid, late-fusion AI framework that combines symptom-based ensemble classification, image-based majority voting, U-Net lesion segmentation, and a clinically guided hybrid severity scoring system aligned with WHO recommendations. By integrating these components, we aim to provide a more comprehensive, interpretable, and practical tool suitable for real-world clinical use, particularly in resource-constrained and outbreak settings.

In conclusion, while previous studies have made important contributions to the classification and detection of Mpox , they lack a hybrid approach that aggregates both clinical symptoms and image-based analysis at the decision level. Our research aims to bridge this gap, offering a more accurate and practical diagnostic tool for the early detection and severity assessment of Mpox.

3. Methodologies

This section outlines the core methodologies adopted in our study. The complete workflow is divided into five key stages: dataset preparation, symptom-based classification, image-based classification, lesion segmentation, and severity scoring.

An overview of these stages is presented in Figure 3.1, depicting the step-by-step progression of our proposed framework. Each stage is briefly introduced here, while detailed explanations that are supported by figures and diagrams, have been provided in the respective sections that follow. This structured approach enables a comprehensive and systematic assessment of Mpox diagnosis and severity estimation.

3.1 Data Collection & Preparation

In this study, multimodal datasets were used to facilitate better classification, segmentation, and severity assessment of Mpox using both clinical symptoms and dermatoscopic images. Since Mpox lesions closely resemble those of other poxviruses, it was essential to take careful and deliberate steps in dataset selection and preparation. Although multiple publicly available datasets exist online, many were found to be unsuitable due to issues such as poor image quality, low resolution, or inconsistent labeling, which could compromise model performance. As a result, we developed a customized dataset tailored to our specific requirements by selectively including or excluding samples as necessary.

Given the hybrid nature of our approach, multiple datasets were utilized across different stages of the methodology. For the image-based classification task, the Monkeypox Skin Images Dataset (MSID), available at https://data.mendeley.com/datasets/r9bfpnvyxr/6 (DOI: 10.17632/r9bfpnvyxr.6), published by Bala and Shamim was used under the Creative Commons Attribution 4.0 International (CC BY 4.0) license [35]. This license permits redistribution and adaptation of the dataset with appropriate credit to the original authors.

A separate dataset containing structured clinical information was employed for symptom-based classification using traditional machine learning algorithms, with license and source details provided in the respective section (3.2) of this paper. Additionally, a distinct and customized dataset was curated for the segmentation task. We derived the segmentation dataset primarily from the MSID dataset, which is covered under the same CC BY 4.0 license, and was further

refined through manual annotation and preprocessing to generate mask files suitable for training the U-Net model.

All datasets were obtained from publicly available and ethically sourced repositories. Necessary modifications and preprocessing steps were performed solely to align them with the specific objectives of each task, in full compliance with their respective license terms. Throughout this manuscript, we use "Mpox" as the standardized disease term in accordance with WHO guidelines, while preserving original dataset names and variable labels where necessary for reproducibility.

3.2 Dataset for Symptom-Based Classification

For symptom-based classification we used Monkey-Pox PATIENTS Dataset, available on Kaggle [36] and modified according to our needs. The Monkey-Pox PATIENTS Dataset is distributed under the CC0: Public Domain license, which permits unrestricted use, distribution, and modification without the need for permission and it is a synthetic dataset designed for training classification models to predict Mpox infection based on clinical symptoms. This synthetic dataset was utilized because publicly available, patient-level Mpox datasets containing structured clinical symptoms, confirmed diagnostic labels, and corresponding lesion images are currently very limited. The dataset was therefore used to develop and test the symptom-processing branch of the proposed framework, rather than to claim clinical validation. It includes structured features like symptom presence and comorbidities, suitable for binary classification tasks (Mpox-positive vs. negative). Table 3.2.1 contains the dataset information.

The Mpox symptom dataset consists of 25,000 patient entries, each represented by 10 features and a target label. The dataset includes clinical symptoms such as Rectal Pain, Sore Throat, Penile Oedema, Oral Lesions, Solitary Lesion, Swollen Tonsils, HIV Infection, and Sexually Transmitted Infection, all encoded as boolean values (True/False). Additionally, the column Systemic Illness contains categorical information about generalized symptoms, though it has missing values in 6,216 entries (non-null count: 18,784) visible on Table 3.2.1.

The target label, MonkeyPox, indicates whether the patient was positive or negative for Mpox. A class distribution plot 3.2.1 reveals a moderate class imbalance, with 15,933 positive cases and 9,067 negative cases. This imbalance was considered in model training using class weights to ensure fair learning. The feature *Patient_ID* was dropped during preprocessing as it does not contribute to the prediction task.

3.2.1 Dataset Processing for Symptom-Based Classification

The symptom dataset was preprocessed by first removing irrelevant identifiers (Patient_ID) and handling missing data, particularly by filling nulls in the Systemic Illness column with "n/a". Boolean features representing clinical symptoms (e.g., rectal pain, sore throat, etc.) were mapped to binary values (0 or 1). The multi-category Systemic Illness field was one-hot encoded to ensure compatibility with machine learning algorithms. An initial exploratory analysis was performed, including dataset statistics, class distribution, and null value checks, to better understand the data before modeling.

A correlation heatmap was generated using all numeric variables, including symptom presence, infection status, and the target variable. This analysis helps visualize the relationships between symptoms and the likelihood of a Mpox diagnosis, visible in Figure 3.2.1.1.

After including the one-hot encoded Systemic Illness categories, a refined heatmap was generated. This visualization (Figure 3.2.1.2) reveals the correlation of individual systemic symptoms with the Mpox outcome, highlighting negative correlations with "n/a" (absence of systemic illness) and mild positive associations with other symptoms.

A correlation matrix was computed between all features and the target variable MonkeyPox. The resulting heatmap (Figure 3.2.1.3) visualizes the strength and direction of these correlations, with deeper red shades indicating stronger positive or negative relationships.

The purpose of this correlation analysis was exploratory, aimed at understanding the linear relationships between individual symptoms and the target variable prior to modeling. As observed in the heatmaps, the correlations between individual symptoms and the Mpox outcome are generally mild (e.g., the strongest positive correlations reaching only approximately 0.14). This indicates that no single symptom serves as a strong standalone linear predictor of the disease. Consequently, explicit feature selection based on these linear correlation thresholds was not performed. Retaining all features was a deliberate choice, as the ensemble of tree-based models utilized in this study (such as LightGBM, XGBoost, and Random Forest) inherently handle irrelevant features and multicollinearity through their node-splitting algorithms without degrading performance. Furthermore, while individual linear correlations are weak, the interaction of multiple mild symptoms may create strong non-linear predictive patterns, which these ensemble models are specifically designed to capture. Thus, the correlation analysis primarily serves to highlight the necessity of using complex, non-linear ensemble methods rather than relying on individual symptom indicators.

3.3 Dataset for Image-Based Classification

As per the license declaration of the dataset, the images were initially gathered from a variety of online sources and were categorized into four classes: Monkeypox, Chickenpox, Measles, and Normal skin. For the purpose of binary classification, all Non-Mpox images (i.e., chickenpox, measles, and normal skin) were consolidated into a single class labeled NonMonkeypox, while Mpox images were kept in a separate folder labeled Monkeypox. This process resulted in a dataset structured into two distinct classes: Monkeypox and NonMonkeypox. Following this classification, the dataset underwent a systematic preparation pipeline: the images were subjected to data augmentation to enhance diversity, the augmented images were saved to a dedicated directory, and finally, a CSV file was generated to map each image file path to its corresponding class label for supervised training.

After applying image augmentation techniques to enhance model generalizability, the final dataset comprised 818 Mpox images and 832 Non-Mpox images. The complete augmented dataset has been made publicly available on Kaggle for transparency and reproducibility [37].

3.3.1 Data Augmentation Strategy for Image-based Classification

To increase the diversity of the training dataset, we employed a comprehensive data augmentation pipeline using the Albumentations library. This strategy helps the model generalize better by simulating various real-world image transformations. Each original image in the dataset was augmented five times using a combination of probabilistic transformations, which included horizontal and vertical flipping, random rotation up to ±40 degrees, and cropping a fixed 200×200 region followed by resizing it to 224×224 pixels to maintain input consistency. Additionally, adjustments to brightness and contrast were randomly applied to mimic variable lighting conditions.

After augmentation, all images were normalized and converted into tensor format for compatibility with deep learning models. This process was applied to every image in the dataset, and the resulting augmented images were saved with distinct filenames in a dedicated directory. Overall, the augmentation process significantly expanded the dataset, enhancing the model's ability to learn invariant features across both Monkeypox and NonMonkeypox classes.

3.3.2 CSV Annotation Generation:

Following the data augmentation phase, we organized the augmented images into two directories: one for Mpox and another for Non-Mpox images. To facilitate supervised learning and streamline the training process, we programmatically generated a CSV file that maps each image file to its corresponding class label. This was achieved using a Python script that traverses both class folders within the dataset directory. It scans for valid image files (with extensions .jpg, .jpeg, or .png) and records the full file path along with the associated label (MonkeyPox or NonMonkeyPox). These records are then structured into a tabular format using Pandas, producing a two-column DataFrame consisting of filepath and label. Finally, this structured dataset is exported as a CSV file named labels_aug.csv.

This CSV serves as a centralized reference for training and evaluating image classification models, ensuring that the file paths and class labels are cleanly aligned and easily accessible during the model training pipeline.

3.3.3 Class Distribution of Image-based Classification

To ensure balanced learning during the training phase of our Mpox classification model, we first examined the distribution of classes in our dataset. The dataset comprises a total of 1,650 images, evenly divided into two categories: MonkeyPox and NonMonkeyPox. Of these, 1,320 images were used for training, while 330 images were reserved for validation. We verified class balance through both a pie chart and a bar plot. The pie chart confirmed that the dataset maintains a nearly equal representation of both classes, minimizing the risk of bias during model training. Additionally, a bar plot was generated to visualize the count of images per class. As shown in Figure 3.3.3.1, both the MonkeyPox and NonMonkeyPox categories contain approximately the same number of samples (over 800 images each), confirming that the dataset is well-balanced. This balance is crucial in binary classification tasks as it helps to prevent the model from favoring one class over the other, leading to more reliable and generalizable performance. A few samples from both of the classes have been provided below in the Figure 3.3.3.2

Here, in the Figure 3.3.3.2, first 4 columns are representing images of Mpox where the last column contains the images classified as Non-Mpox.

3.4 Dataset Preparation for Segmentation

Given that image segmentation is a more complex and fine-grained task compared to classification, meticulous care was taken during the dataset curation process to ensure high-quality inputs for training the segmentation model. For this study, the U-Net architecture was selected owing to its proven performance in biomedical image segmentation tasks and accordingly we tailored the dataset to meet its requirements. Only images captured at close proximity, with clear and visible lesion regions, were included to support precise annotation and model training. This quality control step was crucial to enhance both the accuracy and consistency of the segmentation masks.

Here's a step-by-step diagram (Figure 3.4.1) that visually summarizes our segmentation dataset preparation pipeline, from raw image collection to augmentation and final dataset structure.

For manual annotation, we utilized LabelMe, an open-source image annotation tool known for its flexibility in polygonal segmentation. The tool was installed and run on an Ubuntu 24.04 environment. Each image was annotated manually using polygon outlines around the Mpox lesions, resulting in a corresponding JSON file for every annotated image. To generate the training masks required by our U-Net model, a dedicated Python script was used to convert these JSON files into binary mask images. The resulting masks consisted of white pixel values indicating lesion regions and black pixel values representing the background. This initial phase produced a structured dataset comprising 89 original RGB images (JPG format), 89 JSON annotation files, and 89 binary masks (PNG format). A few sample images selected for segmentation task have been showed in Figure 3.4.2.

3.4.1 Mask Generation from Polygon Annotations

To generate pixel-wise binary segmentation masks from annotated images, we utilized polygon annotations created using the LabelMe tool. Each annotated sample was stored as a JSON file containing polygon coordinates representing lesion boundaries. A custom Python script was developed to systematically convert these annotations into binary mask images in PNG format. Initially, all JSON files were collected from the designated annotation directory. For each file, the labelme_json_to_dataset utility from the LabelMe toolkit was executed to extract the corresponding label image (label.png), which encodes class information as pixel intensities. This raw mask was then processed using the Python Imaging Library (PIL) and NumPy to ensure binary format consistency like pixels corresponding to lesions were assigned a value of 255 (white), and the background was set to 0 (black). This binarization step was crucial for compatibility with downstream segmentation models, particularly U-Net and Mask R-CNN architectures that require masks in binary format. Finally, the processed binary masks were saved with the same base name as their corresponding JSON files and stored in a dedicated output directory. Temporary dataset folders generated during conversion were automatically deleted to maintain a clean working directory. This preprocessing ensured that all images, masks, and

annotations were correctly aligned and ready for subsequent data augmentation and model training stages.

3.4.2 Data Augmentation and Annotation Transformation

To enhance the generalizability of our model, we applied comprehensive data augmentation techniques on the original dataset. The augmentation pipeline was constructed using the Albumentations library, incorporating spatial and photometric transformations such as horizontal and vertical flips (p=0.5), random rotations up to 40 degrees (p=0.7), and random-sized cropping to simulate real-world variability. Additionally, random brightness and contrast adjustments were introduced (p=0.2) to account for different imaging conditions. All images were normalized and converted to PyTorch tensors using ToTensorV2() to ensure compatibility with the training pipeline.

The augmentation process was not limited to image and mask pairs but was extended to the corresponding polygon annotations provided in the LabelMe JSON format. To preserve the spatial integrity of lesion annotations during augmentation, each polygon was projected onto a binary mask and the same augmentation transformations were applied synchronously to both the image and its binary mask. Contours were then re-extracted from the transformed mask to generate updated polygon points. This procedure ensured accurate transformation of lesion boundaries while maintaining semantic consistency.

Each image in the dataset was augmented five times, resulting in a fivefold (89 x 5 = 445) increase in dataset size. For every augmented instance, the corresponding image, binary segmentation mask, and updated annotation file were saved in dedicated directories. The updated JSON files preserved the original metadata while incorporating the new polygon coordinates and image paths, ensuring seamless integration into downstream segmentation pipelines. This augmentation strategy not only diversified the dataset but also preserved the semantic fidelity of annotations, which is crucial for training deep learning models on tasks requiring high spatial precision, such as lesion segmentation in medical images.

To support transparency, reproducibility, and further research, we have made the full dataset publicly available on Kaggle [38], including both the original and augmented files across all formats (JPG, JSON, and PNG).

3.4.3 Visualization After Augmentation

This stage is essential to ensure the integrity and consistency of the augmented dataset before training the proposed U-Net model. During the augmentation process, not only the original images but also their corresponding JSON annotation files and ground truth masks were augmented. Therefore, it is critical to verify that the augmented images and their respective masks remain properly aligned. This step serves as a validation checkpoint to confirm that the dataset is stable and accurately prepared for the subsequent training phase. Figure 3.4.3.1 illustrates the original images alongside their corresponding masks before and after augmentation, clearly demonstrating that the augmented masks are accurately aligned with the augmented images. In the last row, we randomly picked an image and checked its corresponding mask.

After that, we checked for duplicated of inconsistent masks but found none. Before training, the dataset was evaluated for class imbalance and shape consistency too. Where lesion pixels accounted for 6.81% of total mask pixels. Additionally, all images and masks were confirmed to have consistent shapes (224, 224, 3) for images and (224, 224, 2) for one-hot encoded masks for ensuring compatibility with the U-Net architecture.

3.4.3.1 Pixel Intensity Distribution Analysis

Prior to conducting segmentation-specific visualizations, we examined the overall pixel intensity distribution across the training dataset to gain insights into the dynamic range and tonal composition of the input images. A histogram of pixel values was generated using the entire training set (Figure 3.4.3.1.1), where pixel intensities were normalized to the range [0, 1] and aggregated into 50 bins.

The distribution reveals a quasi-normal profile, with a gradual increase in frequency from lower intensities and a peak density occurring around mid-range values (0.5–0.7). The curve then gently tapers off toward higher intensities, with several distinct local maxima. These peaks may correspond to common background tones, skin textures, or lighting conditions in the dermatoscopic or clinical image acquisition settings. Notably, the presence of intensity values spanning almost the entire normalized range suggests that the dataset possesses sufficient contrast and variability, which is favorable for feature learning by convolutional neural networks (CNNs). Moreover, the smoothness of the histogram and absence of abrupt spikes or gaps indicate a relatively uniform distribution of intensities, further supporting the effectiveness of the image augmentation and preprocessing pipeline.

This analysis confirms that no significant clipping or saturation is present in the pixel value space, thereby validating the dataset's readiness for downstream tasks such as segmentation.

3.4.3.2 Sample-Based Histogram Analysis of Image and Mask Intensities

To further validate the characteristics of the dataset at the individual image level, we conducted a histogram analysis on a randomly selected augmented image and its corresponding binary segmentation mask (Figure 3.4.3.2.1). This allows us to examine local intensity behavior and ensure that the image–mask pairs are well-aligned and correctly preprocessed.

i. **Left Panel – Image Histogram:** The grayscale image shows a multi-modal intensity distribution, primarily concentrated in the range of 140–230. This suggests the presence of multiple tissue types or lighting variations within the image. The peaks around intensity levels 150 and 210 may correspond to different skin textures or lesion regions and surrounding healthy tissue. The absence of values near the extremes (i.e., close to 0 or 255) indicates a balanced exposure, minimizing the risk of clipping or saturation.

ii. **Right Panel – Mask Histogram:** The mask histogram reveals a highly bimodal distribution, with dominant peaks near 0 and 255, which corresponds to background and lesion regions, respectively. This binary nature confirms that the mask has been correctly encoded and retains its integrity as a ground truth segmentation label. The significantly higher frequency of the 0-intensity pixels highlights the sparse nature of lesion regions relative to the background, which is typical in medical segmentation tasks.

This visual validation confirms the correctness of grayscale loading, the expected binary behavior of masks, and the preservation of contrast in images. It further supports the suitability of the dataset for deep learning-based segmentation models, which benefit from both intensity diversity in the images and clarity in the ground truth labels.

3.4.3.3 Pairwise Intensity Distribution Analysis Between Images and Segmentation Masks

To investigate the pixel-wise relationship between the original skin lesion images and their corresponding segmentation masks, we implemented a custom visualization pipeline using Python libraries such as OpenCV, NumPy, Seaborn, and Matplotlib. The objective was to quantitatively and visually assess the statistical relationship between pixel intensity distributions, which is crucial in validating the quality and relevance of the segmentation masks used for supervised learning.

In our pipeline, grayscale pixel intensities were extracted from a random sample of image-mask pairs from the augmented dataset. Each image and its corresponding binary segmentation mask were flattened into one-dimensional arrays, and their pixel intensities were stored in a structured DataFrame. This dataset was then used to generate an enhanced pair plot (Figure 3.4.3.3.1), which consists of marginal kernel density estimations (KDE) along the diagonals and joint KDE scatter plots in the lower triangle.

1. **Marginal Distributions (Diagonal):**
    a. The left diagonal panel shows the distribution of pixel intensities in the original images, which are relatively continuous and normally distributed, spanning a wide range from low to high intensity.
    b. The right diagonal panel illustrates the distribution of mask intensities, which are highly discrete and bimodal, with values clustering near 0 and 255. This is expected since segmentation masks are binary, representing lesion (foreground) and background.
2. **Joint Distributions (Lower Triangle):**
    a. The bottom-left joint KDE plot visualizes the joint distribution between image intensities and mask intensities.
    b. The observed pattern indicates that there is minimal correlation between the raw pixel intensity of the input images and the pixel intensity of the masks. This is further supported by the computed Pearson correlation coefficient of 0.03, as shown in red.

This low correlation implies that the mask delineation is not trivially inferred from raw intensity values alone, suggesting the presence of complex spatial and textural features defining the lesions. Consequently, deep learning models, particularly convolutional neural networks (CNNs) are well-suited for capturing such non-linear and spatial dependencies in the segmentation task.

3.5 Symptom-Based Classification Using Ensemble Learning

To develop a reliable classification model capable of identifying Mpox infections based solely on symptomatology, we implemented a multi-model ensemble learning pipeline. The input

dataset comprised structured clinical symptom data, where the target variable was binary (Mpox : 1 for positive, 0 for negative). Initially, the dataset was split into independent variables (features) and the target label. We employed an 80/20 train-test split strategy using a fixed random seed (42) to ensure reproducibility.

3.5.1 Handling Class Imbalance

Although the dataset contains a greater number of positive (Mpox) cases (15,933) compared to negative cases (9,067), the class distribution remains asymmetrical, pointing adjustment to prevent biased learning. To address this, class-specific weights were computed using scikit-learn's compute_class_weight method with the 'balanced' strategy. These weights were incorporated into all compatible classifiers, such as XGBoost, CatBoost, and LightGBM, to ensure equitable learning and generalization across both classes.

3.5.2 Model Selection and Hyperparameter Optimization

To leverage the predictive strengths of diverse machine learning paradigms, six classifiers were selected as base learners: Light Gradient Boosting Machine (LightGBM), eXtreme Gradient Boosting (XGBoost), CatBoost, Logistic Regression (LR), Random Forest (RF), and Support Vector Classifier (SVC). These classifiers were chosen for their complementary properties, ranging from linear separability and ensemble diversity to efficient handling of high-dimensional feature spaces.

For the gradient boosting models like LightGBM, XGBoost, and CatBoost, a comprehensive hyperparameter tuning was conducted to optimize generalization capability while ensuring computational feasibility on CPU-based environments. Optimization focused on key parameters such as learning rate, maximum tree depth, regularization penalties (reg_alpha, reg_lambda), number of estimators, and subsampling ratios. This tuning process was performed empirically with the goal of balancing bias-variance trade-offs and improving model's performance.

1. **LightGBM** was configured with boosting_type='gbdt', a learning rate of 0.0165, maximum depth of 3, and L1/L2 regularization (reg_alpha=7.555, reg_lambda=0.953).
2. **XGBoost** adopted a histogram-based optimization strategy via tree_method='hist', with an adaptive learning rate of 0.0157 and class-balancing via scale_pos_weight.
3. **CatBoost** employed bootstrap_type='Bernoulli' and grow_policy='Depthwise', with a learning rate of 0.0175 and tree depth of 3.

Table 3.5.2.1 contains more detailed information on these models.

All models were executed using CPU backends to promote reproducibility and accessibility without specialized hardware requirements. The resulting classifiers were each independently assessed using stratified cross-validation to ensure fairness and generalizability.

3.5.3 Hierarchical Stacking Ensemble Architecture

To consolidate the predictive strengths of the individual classifiers, a three-tier stacking ensemble was constructed. This architecture is illustrated in Figure 3.5.3.1 and consists of two meta-learning stages designed to enable deep hierarchical integration of model predictions.

Here in the Figure 3.5.3.1, the base learners consist of six classifiers feeding into a two-stage meta-learning process. The first meta-layer integrates outputs from gradient boosting models, while the final meta-learner combines all components to produce the final prediction.

1. **Level-0: Base Learners**

The initial layer consists of six heterogeneous base classifiers: LightGBM, XGBoost, CatBoost, Logistic Regression, Random Forest, and SVC. Each model was trained independently and output class probability estimates. These outputs serve as a diverse feature set for the subsequent meta-learners.

2. **Level-1: Gradient Boosting Sub-Ensemble**

The second layer functions as an intermediate ensemble that aggregates the outputs of the three gradient boosting models (LightGBM, XGBoost, and CatBoost). A logistic regression model was trained as a meta-learner on these outputs to capture non-linear interactions and complementarity among the tree-based classifiers.

3. **Level-2: Final Ensemble Estimator**

The final predictive layer integrates the complete set of Level-0 base learners and the Level-1 boosting ensemble. These are combined through a top-level logistic regression classifier, which serves as the final decision function. This recursive ensemble architecture promotes hierarchical learning of feature interactions, enabling the model to capture both low-level and high-level patterns across model families.

To ensure the prevention of information leakage across layers, three-fold cross-validation (CV=3) was employed during the stacking process. This approach mitigates the risk of overfitting and enhances the generalization performance of the final meta-model.

3.5.4 Evaluation Metrics and Model Performance

In this part, the performance of the proposed symptom-based binary classification model for Mpox prediction was rigorously evaluated using multiple quantitative metrics. These metrics provide comprehensive insights into the model's diagnostic reliability and generalization capacity across both training and unseen test data. Specifically, accuracy, confusion matrices, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) were employed to assess model effectiveness.

3.5.4.1 Confusion Matrix Analysis

The confusion matrices presented in Figure 3.5.4.1.1 illustrate the distribution of predicted versus actual classes on both training and test datasets, enabling interpretability beyond overall accuracy. The corresponding accuracy metrics are visually illustrated in Figure 3.5.4.1.1.

The confusion matrices show that the symptom-based model is more effective at identifying mpox-positive cases than mpox-negative cases. False positives outnumber false negatives, indicating that the classifier is biased toward predicting the positive class. While such a sensitivity-oriented behavior may be useful for preliminary screening, it also means that a considerable proportion of Non-Mpox cases may be incorrectly flagged as suspected mpox. Therefore, the symptom-based classifier should not be interpreted as a standalone diagnostic

tool. Instead, it should be considered a preliminary risk-stratification component that requires confirmation through lesion image analysis, clinician assessment, and/or laboratory testing.

3.5.4.2 Precision, Recall, and F1-Score

To obtain a more granular understanding of class-wise model behavior, precision, recall, and F1-score were analyzed for both training and test datasets, as summarized in Table 3.5.4.2.1. These metrics are particularly informative in imbalanced classification settings, where accuracy alone may obscure clinically relevant performance trends.

The model demonstrates high recall for mpox-positive cases, with class 1 recall of approximately 0.89 across training and test sets. This indicates that the classifier is sensitive to suspected mpox cases and produces relatively few false negatives. However, recall for the negative class remains low, approximately 0.35–0.37, indicating limited ability to correctly identify Non-Mpox cases. This low negative-class recall reflects reduced specificity and a tendency toward false-positive predictions. In practical settings, such behavior could lead to unnecessary concern, additional confirmatory testing, and possible over-triage of Non-Mpox patients. Therefore, the symptom-based classifier is best interpreted as a supportive screening or risk-stratification module rather than a definitive diagnostic model. The trade-off between positive-case sensitivity and negative-case specificity should be calibrated according to the intended use case, and importantly, the close alignment of training and test metrics suggests stable generalization without evidence of overfitting.

3.5.4.3 ROC Curve and AUC Analysis

The Receiver Operating Characteristic (ROC) curves for both the training and test sets are depicted in Figure 3.5.4.3.1, alongside the corresponding Area Under the Curve (AUC) values:

- AUC (Training Set): 0.70
- AUC (Test Set): 0.71

The close alignment of AUC values between the training and test sets suggests that the model does not exhibit substantial overfitting within the internal split. However, the absolute AUC value indicates only moderate discriminative ability. Therefore, the symptom-based classifier should not be considered clinically reliable as an independent diagnostic model. Instead, its performance supports its use as an auxiliary component within the proposed late-fusion framework. The moderate AUC also reflects the inherent limitations of symptom-only prediction, particularly because mpox symptoms overlap with other viral, dermatological, and sexually transmitted infections.

From a clinical perspective, the model exhibits a high recall for class 1 (infected patients), reflecting a strong bias toward sensitivity. While this is advantageous in early detection scenarios where minimizing false negatives is essential to ensure timely intervention and prevent further transmission, this bias must be balanced against the high false-positive rate indicated by the low negative-class recall. Overall, the model's current performance provides a valuable baseline for automated mpox risk stratification based on symptom data alone, but future work should focus

on validating and recalibrating the symptom classifier using real patient-level clinical records with confirmed diagnostic outcomes.

3.6 Design and Optimization of Deep CNN Models for Skin Image-Based Mpox Classification

To address the task of automated Mpox image classification, we employed transfer learning using seven Convolutional Neural Network (CNN) architectures: DenseNet121, EfficientNetB0, InceptionV3, MobileNetV2, VGG16, ResNet50, and Xception. These models were selected based on their proven performance across medical image analysis benchmarks, architectural diversity, and varying computational complexities, thereby enabling a holistic evaluation across different learning paradigms.

All selected models were pre-trained on the ImageNet dataset, which contains over 14 million labeled images across 1000 categories. The pre-trained weights provided these networks with the ability to extract general-purpose visual features such as edges, shapes, and textures that are transferable across domains, including medical imaging. By leveraging these transferable features and fine-tuning the networks on the Mpox dataset, we aimed to minimize training time, reduce the risk of overfitting, and improve performance despite the relatively limited size of medical image datasets.

Each model was adapted by removing the original classification head and appending a custom classifier block, designed to suit the binary classification nature of our task (Mpox vs. Non-Mpox). Furthermore, to standardize input dimensions across all models, all images were resized to 224×224 pixels with three color channels (RGB), which aligns with the input expectations of the pre-trained backbones.

Figure 3.6.1 shows the transfer learning framework used for Mpox image classification using seven pre-trained CNN models: DenseNet121, EfficientNetB0, InceptionV3, MobileNetV2, VGG16, ResNet50, and Xception. Input images resized to 224×224×3 are fed into these models. where the early convolutional blocks are frozen to retain generic visual features, while the deeper convolutional blocks are unfrozen and fine-tuned for Mpox-specific patterns. The original top layers are replaced with a custom head (e.g., GlobalAveragePooling2D, Dense, Dropout, and a Softmax output) for binary classification. The models then output a probability indicating the presence or absence of Mpox.

3.6.1 Architectural Details and Training Strategies

I. **DenseNet121** [59]

a. In our implementation, the first 300 layers were frozen to retain generic visual features, while the remaining 111 layers were unfrozen and fine-tuned.

b. The custom classifier included a Global Average Pooling (GAP) layer to reduce dimensionality, followed by Batch Normalization and Dropout (rate: 0.5) to prevent overfitting. A Dense layer with 512 neurons and ReLU activation was used for high-level abstraction, followed by another Batch Normalization and a Dense output layer with softmax activation to predict the two classes.

Training used the Adam optimizer with a learning rate of 0.0001, chosen for its adaptive gradient estimation, and was run for 40 epochs with EarlyStopping and ReduceLROnPlateau callbacks to avoid overfitting and optimize learning dynamics.

II. **EfficientNetB0** [60]

a. We froze the entire base model to retain pre-trained features and avoid overfitting given the moderate dataset size.
b. A streamlined classification block was added: GAP → Dropout (0.5) → Dense (512 neurons, ReLU) → Dense (2 neurons, softmax).

Training employed Stochastic Gradient Descent (SGD) with a learning rate of 0.0009 and momentum (default: 0.9), over 45 epochs. This setup favors gradual convergence and generalization when fine-tuning shallow classifiers on top of a fixed backbone.

III. **InceptionV3** [61]

a. We opted to freeze all layers except the last 30, allowing selective fine-tuning of deeper, high-level abstract features specific to skin lesion identification.
b. The classifier comprised a GAP layer, followed by Batch Normalization, Dropout (0.5), a Dense layer (512 neurons, ReLU), a second Batch Normalization, and a Dropout (0.4) before the final Dense softmax layer.

The model was trained using Adam with a learning rate of 0.0001 for 40 epochs, again leveraging learning rate scheduling and early stopping to ensure convergence.

IV. **MobileNetV2** [62]

a. Due to the simplicity of the network and the risk of overfitting on a small dataset, we froze all layers of the base model.
b. A custom head was added: Flatten → Dropout (0.5) → Dense (512, ReLU) → Dense (2, softmax).

This model was trained for 50 epochs using Adam with a learning rate of 0.0001. Despite its simplicity, MobileNetV2 demonstrated promising performance with minimal computational cost.

V. **VGG16** [63]

a. We froze the entire pre-trained model, as preliminary experiments showed that unfrozen layers tended to overfit quickly.
b. The classifier comprised a Flatten layer, a Dense layer (512 neurons, ReLU), and a Dense output layer with softmax activation.

The model was trained using Adam (learning rate: 0.0001) for 35 epochs. Although no dropout or batch normalization was applied, the model maintained stable performance due to the frozen base.

VI. **ResNet50** [64]

a. The entire base model was frozen, allowing only the classifier to learn.
b. The appended head included: GAP → Dropout (0.5) → Dense (512, ReLU) → Dense (2, softmax).

Training was performed using SGD (learning rate: 0.0009) over 50 epochs, aiming for stable, regularized updates in the fixed-feature setting.

VII. Xception [65]

a. In our setup, only the final 5 layers were unfrozen, facilitating minor domain-specific fine-tuning while preserving the general feature extractor.
b. The classifier included: GAP → Batch Normalization → Dropout (0.5) → Dense (256 neurons, L2 regularized) → Dropout (0.4) → Dense (2 neurons, softmax, L2 regularized).

This model was trained with the Adam optimizer at a learning rate of 0.0005 for 40 epochs, using L2 regularization ($\lambda = 0.01$) to minimize overfitting risk in the deeper layers.

Justification for Heterogeneous Classification Heads and Training Configurations: Rather than applying a uniform classification head across all models, the architectures and training configurations were individually tailored to balance each model's capacity, trainable parameter count, and risk of overfitting. Models with unfrozen base layers (e.g., DenseNet121, InceptionV3, Xception) possess a significantly higher number of trainable parameters and a greater risk of overfitting on a limited dataset; consequently, they required stronger regularization mechanisms, such as dual Batch Normalization layers, multiple Dropout blocks, and L2 weight penalties. Conversely, models with completely frozen backbones (e.g., EfficientNetB0, ResNet50) have far fewer trainable parameters confined to the classifier head, allowing for simpler architectures with single Dropout layers. Furthermore, the choice between Global Average Pooling (GAP) and Flatten layers was dictated by the output shape and feature density of the respective backbones. For instance, VGG16 and MobileNetV2 naturally yield dense feature maps that pair effectively with a Flatten layer, while deeper, more complex architectures benefit from GAP to reduce dimensionality prior to classification. This customized, model-specific approach ensures that each network achieves optimal convergence and generalization without succumbing to overfitting or underfitting.

The summary of these 7 CNN models have been given below on Table 3.6.1.1.

All models were compiled with the sparse categorical crossentropy loss function, suitable for integer-labeled binary classification tasks. Each model's training configuration was customized based on its architectural depth, feature output dimensions, and transferability characteristics, as detailed in Table 3.6.1.1 and justified above.

3.6.2 Evaluation Metrics and Model Performance

To assess the performance of the seven pre-trained CNN models fine-tuned for Mpox image classification, we evaluated both training dynamics and final predictive outcomes using accuracy and loss metrics. The training and validation accuracy and loss curves for each model are presented in Figure 3.6.2.1, which offers a visual insight into the learning behavior and convergence over epochs.

The evaluation of model performance was based on key metrics such as validation accuracy and validation loss. These values are summarized in Table 3.6.2.1: Accuracy and Loss Summary, while the corresponding training and validation curves are illustrated in Figure 3.6.2.1, showcasing the learning behavior across epochs for each of the seven models.

DenseNet121 demonstrated strong and stable learning, achieving a high validation accuracy of 96.36% with a low validation loss of 9.18%. This reflects both effective feature extraction and strong generalization capabilities. Similarly, ResNet50 also attained 96.67% accuracy and 9.67% loss, suggesting better performance and convergence. Among all models, VGG16 outperformed the rest, reaching the highest validation accuracy of 97.58% while maintaining a low validation loss of 10.74%, despite its relatively older architectural design. EfficientNetB0, though slightly trailing the top performers, still achieved an impressive 94.55% accuracy and 17.27% validation loss, highlighting its capacity for efficient learning. MobileNetV2 delivered competitive results, with a 91.82% accuracy and 20.36% loss, indicating strong performance with a lightweight architecture. On the other hand, InceptionV3 and Xception exhibited relatively lower performance, with accuracies of 89.09% and 91.52%, and higher losses of 43.49% and 39.43%, respectively. These findings suggest potential issues such as overfitting or suboptimal fine-tuning that may have hindered their full adaptation to the Mpox image dataset.

Overall, the results clearly indicate that DenseNet121, ResNet50, and VGG16 were the most effective models for this classification task, achieving an optimal balance between accuracy and loss. These outcomes are further supported by their smooth convergence curves and minimal discrepancy between training and validation metrics as observed in the corresponding graphs.

3.7 Mpox Lesion Segmentation using Customized U-Net Architecture

In this part, we utilized a customized version of the U-Net architecture, originally proposed by Ronneberger et al. (2015) for biomedical image segmentation [66]. Our implementation expands on the original design by incorporating additional convolutional blocks and deeper layers to improve feature abstraction and localization on complex lesion patterns. Figure 3.7.1 illustrates the overall U-Net structure used in our model.

3.7.1 Architectural Design of the U-Net Model

The proposed customized architecture follows the standard encoder–decoder paradigm with skip connections, consisting of the following components:

i. **Input Layer:** The model accepts input images of shape (224 × 224 × 3), corresponding to RGB dermoscopic or clinical images of Mpox lesions resized during preprocessing.

ii. **Encoder (Contracting Path):** The encoder comprises four levels of downsampling. Each level contains two convolutional layers followed by a max-pooling operation. The number of filters increases progressively (64, 128, 256, 512), allowing the network to learn hierarchical features of increasing complexity. The convolutional layers employ 3×3 kernels, ReLU activation, and 'same' padding to maintain spatial dimensions. A fifth encoder level was added to further increase the receptive field. This level consists of two 512-filter convolutions and a max-pooling operation, enabling deeper feature abstraction.

iii. **Bridge (Bottleneck):** The bridge between the encoder and decoder paths consists of two convolutional layers with 1024 filters. These layers process the most abstracted features and encode high-level semantic information about the lesion regions.

iv. **Decoder (Expanding Path**): The decoder path mirrors the encoder with four levels of upsampling using transposed convolutions (stride = 2) followed by concatenation with

corresponding encoder feature maps via skip connections. These skip connections mitigate the loss of spatial information by reintroducing fine-grained details from the encoder. After each concatenation, two 3×3 convolutional layers are applied to refine the upsampled feature maps.

v. **Output Layer:** The final layer is a 1×1 convolution with sigmoid activation, producing a binary mask where each pixel value represents the probability of lesion presence. The output has the same spatial dimensions as the input (224×224×1), suitable for pixel-wise binary classification.

This extended U-Net architecture leverages deep spatial and contextual feature integration, making it more suitable against lesion variability in real-world clinical scenarios.

3.7.2 Custom Evaluation Metrics for Mpox Lesion Segmentation

In the context of medical image segmentation, standard evaluation metrics such as accuracy are often insufficient, especially when dealing with imbalanced datasets where the lesion area occupies only a small fraction of the total image. To better assess the performance of the proposed U-Net model for Mpox lesion segmentation, we implemented two specialized metrics: the Dice Coefficient and the Intersection over Union (IoU), both customized as differentiable TensorFlow functions.

1. **Dice Coefficient (F1 Score):** The Dice coefficient is defined as:

$$Dice(G,P) = \frac{2.|G \cap P| + \epsilon}{|G| + |P| + \epsilon} \qquad (1)$$

Where, |G∩P| is the number of pixels common to both the ground truth and the predicted segmentation (i.e., the intersection), |G| and |P| are the total number of pixels in the ground truth and predicted masks respectively, ϵ is a small constant ((e.g., 1×10^{-6}) added for numerical stability to avoid division by zero.

The Dice coefficient ranges from 0 to 1, where a value of 1 indicates perfect agreement between the predicted and ground truth masks, and 0 indicates no overlap. Due to its F1-score-like formulation, it balances both false positives and false negatives, making it particularly effective for segmenting small lesion regions in large medical images.

2. **Intersection over Union (IoU):** It is defined as:

$$IoU(G,P) = \frac{|\ G \cap P\ | + \epsilon}{|\ G \cup P\ | + \epsilon} = \frac{|\ G \cap P\ | + \epsilon}{|G| + |P| - |G \cap P| + \epsilon} \qquad (2)$$

Where, |G∩P| denotes the number of correctly predicted positive pixels (true positives), |G∪P| represents the total number of pixels in either the predicted or ground truth masks (union), ϵ is a small constant.

These metrics are preferred over standard accuracy for medical segmentation due to severe class imbalance.

3.7.2.1 Training Config

The dataset was preprocessed and split into training and validation sets using an 80/20 split. Data was loaded using a custom function load_data(), which prepares image-mask pairs for training.

Training was conducted using the Adam optimizer with a learning rate of $1\times10-4$, a batch size of 4, and for 50 epochs. To ensure optimal model performance and prevent overfitting, a ModelCheckpoint callback was employed to save the best model based on validation loss during training:

3.7.3 Qualitative Segmentation Results

To visually assess the segmentation performance of the proposed U-Net model, representative samples of test images, ground truth annotations, and predicted lesion masks are illustrated in Figure 3.7.3.1. The figure is divided into two panels for enhanced visualization. The left panel presents results in grayscale colormap (cmap=gray), where each row includes the original Mpox image, the corresponding binary ground truth mask, and the predicted mask generated by the model. The two rightmost columns present the corresponding samples rendered with a perceptual colormap, in which lesions are emphasized in yellow against a purple background to enhance contrast and interpretability.

As observed, the predicted masks exhibit a strong spatial correspondence with the annotated ground truth across various lesion shapes, sizes, and skin textures. The model effectively captures both single and multiple lesion regions, demonstrating its ability in handling the morphological variability inherent in real-world Mpox presentations. Slight boundary differences, visible in the colored predictions, are likely due to soft activations at the edges and the post-processing threshold applied to convert probability maps into binary masks. These visual results further substantiate the quantitative performance metrics (Dice and IoU) discussed earlier, confirming that the U-Net model can accurately and consistently segment Mpox lesions across diverse patient samples. The segmentation masks produced are reliable enough to be used in downstream tasks such as severity scoring and lesion counting of the proposed framework

3.8 Hybrid Severity Assessment Based on Clinical Symptoms and Image Segmentation Outputs

At this stage, we designed a comprehensive severity calculation framework that combines symptom-based assessment with image-based lesion analysis to provide a holistic evaluation of Mpox infection severity. This hybrid, late-fusion approach ensures that both clinical manifestations and visual disease burden are captured in the final severity score.

Figure 3.8.1 illustrates the step-by-step workflow of the proposed severity assessment framework for Mpox patients, combining both symptom-based evaluation and lesion-based image analysis. The workflow begins with input data acquisition, where clinical symptoms and skin lesion images are collected. The lesion images are processed through a pre-trained U-Net segmentation model to predict lesion masks. Lesion characteristics such as number of lesions and total lesion area are extracted. Simultaneously, the clinical symptoms are used to calculate a symptom severity score based on predefined symptom weights, guided by the World Health Organization's (WHO) smallpox eradication program guidelines [39] [40] [41] [42]. On the other hand, if either the symptom assessment, the image analysis, or both are classified as negative, the related variables are set to be non-effective in the severity calculation. Both the symptom score and the image score are normalized and then combined using a weighted

summation approach to compute the final severity score. Based on the final score and defined thresholds (T1, T2, T3), the patient's condition is categorized into one of four severity levels: Mild, Moderate, Severe, or Serious. This integrated framework enables a more comprehensive and clinically aligned severity stratification for effective patient triage and management.

3.8.1 Symptom-Based Severity Score

The first component of the framework focuses on patient-reported symptoms. A set of twelve key symptoms commonly associated with mpox infections was selected, including "Rectal Pain," "Sore Throat," "Penile Oedema," "Oral Lesions," "Swollen Tonsils," "HIV Infection," and manifestations of systemic illness such as fever and muscle aches. Each symptom was assigned a specific weight based on its clinical significance, with higher weights attributed to more severe or critical symptoms. For instance, "HIV Infection" and "Oral Lesions" were assigned greater weights compared to minor symptoms. The assigned values are given in Table 3.8.1.1.

Each symptom is assigned a specific weight ranging from 0 to 15, reflecting its relative clinical risk and prognostic significance in Mpox severity assessment. Higher weights indicate a stronger contribution to overall disease severity. HIV infection is given the maximum weight of 15 due to the substantially increased risk of complications and mortality in immunocompromised patients, as emphasized by WHO and CDC guidelines. Oral lesions and swollen lymph nodes, each assigned a weight of 12, signal systemic or mucosal involvement and are often associated with painful, complicated clinical courses. Systemic fever, sexually transmitted infections (STIs), and rectal pain, weighted at 10, reflect aggressive inflammatory responses and a higher risk of severe outcomes such as dehydration or secondary infections. Penile oedema and muscle aches are moderately severe indicators with weights of 8, representing local or systemic viral spread. Solitary lesions and swollen tonsils are weighted between 6 and 7, indicating less severe but clinically important findings. Finally, sore throat (weight = 5) which is considered mild. We are taking only these inputs since our model was specifically trained based on these parameters.

For each patient, the symptom profile was encoded in a binary form (presence or absence), and the symptom severity score $S_{symptoms}$ was calculated as follows:

$$S_{symptom} = 100 \times \left(\frac{\sum_{i=1}^{n} w_i \times p_i}{\sum_{i=1}^{n} w_i} \right) \qquad (3)$$

Where, W_i is the weight assigned to the i-th symptom, and P_i denotes symptom presence (1) or absence (0). This method enabled a normalized score out of 100, allowing for direct integration with image-based metrics.

3.8.2 Image-Based Severity Score

The image-based analysis utilized the pre-trained U-Net model to localize and quantify mpox skin lesions. Each input image was resized to match the U-Net input dimensions and processed to generate a probability mask for lesion regions. The predicted mask was then thresholded to obtain a binary lesion mask. Connected component analysis was subsequently applied to the binary mask to extract two lesion-burden metrics: total lesion area, A, defined as the sum of all segmented lesion pixels, and number of discrete lesions, N, defined as the number of connected lesion components.

The lesion-count metric was normalized using a predefined upper normalization bound, N_{max}. In this study, N_{max} was set to 50 as a framework-specific upper bound for lesion-count normalization. This value was selected to represent a high lesion burden within the proposed scoring framework and to prevent the lesion-count component from increasing without bound. However, N_{max} should not be interpreted as a clinically validated diagnostic threshold; rather, it is a normalization parameter used within the proposed prototype severity-scoring system.

The lesion-area metric was normalized using the total image area rather than an arbitrary maximum pixel value. For an input image of size H × W, the lesion area ratio is defined as A / (H × W), where A is the total segmented lesion area in pixels and H × W represents the total number of pixels in the resized input image. Since all images in this study were resized to 224 × 224 pixels before processing, the total image area was 50,176 pixels.

The image-based severity score, S_{image}, was calculated as:

$$S_{image} = 100 \times \left(\alpha \,.\min\left(\frac{N}{N_{max}}, 1 \right) + \beta.\min\left(\frac{A}{H \times W}, 1 \right) \right) \quad (4)$$

where N is the number of detected lesions, N_{max} is the lesion-count normalization bound, A is the total segmented lesion area, and H × W is the total image area. The min function caps each normalized component at 1 to ensure that the final image-based severity score remains bounded. In this study, α was set to 0.6 and β was set to 0.4. Greater weight was assigned to lesion count because the number of discrete lesions provides an interpretable indication of lesion burden. Lesion area was assigned a lower weight because large coalescing lesions may not always indicate greater clinical severity than multiple smaller discrete lesions.

This image-based severity score is intended as a normalized lesion-burden indicator within the proposed framework. It should not be interpreted as a clinically validated standalone severity index. Future work should calibrate the lesion-count and lesion-area weighting parameters using real patient-level datasets with clinician-assigned severity labels and clinical outcomes.

3.8.3 Final Score Aggregation and Thresholds

After obtaining the symptom-based severity score and the image-based severity score, the two values were combined using a late-fusion weighted aggregation strategy. The final severity score was calculated as:

$$S_{final} = \min(\gamma.S_{symptom} + \delta.S_{image}, 100) \quad (5)$$

where $S_{symptom}$ represents the normalized symptom-based severity score, Simage represents the normalized image-based severity score, and γ and δ represent the fusion weights assigned to the symptom and image branches, respectively. In this study, both γ and δ were set to 0.5 to give equal contribution to the two modalities in the final score.

The final score was then mapped into four ordinal severity categories using the following rule:

If $S_{final} < T1$, the case is classified as Mild.

If $T1 \leq S_{final} < T2$, the case is classified as Moderate.

If $T2 \leq S_{final} < T3$, the case is classified as Severe.

If $S_{final} \geq T3$, the case is classified as Serious.

Where T1, T2, and T3 are framework-specific cutoffs applied to the normalized final severity score, not raw lesion-count thresholds. In this pilot framework, these cutoffs were set to 20, 50, and 75, respectively, to divide the 0–100 composite severity score into mild, moderate, severe, and serious categories.

3.8.4 Threshold Rationale and Sensitivity Analysis

The final severity score is a normalized composite index ranging from 0 to 100. Therefore, the thresholds T1 = 20, T2 = 50, and T3 = 75 are not direct diagnostic thresholds or raw lesion-count thresholds. Instead, they are framework-specific pilot cutoffs used to map the normalized final score into four ordinal severity categories: mild, moderate, severe, and serious. These cutoffs were selected to provide an interpretable initial stratification of increasing clinical concern, where higher scores reflect greater combined symptom burden and lesion burden.

We acknowledge that these thresholds are deterministic and guideline-informed rather than empirically optimized. Clinical severity in mpox may depend on multiple factors, including lesion burden, symptom severity, lesion location, systemic complications, immune status, pain, and disease progression. Therefore, lesion count or a composite score alone is insufficient to define clinical severity. For this reason, the proposed thresholds should be interpreted as preliminary framework-specific score bands rather than clinically validated severity cutoffs.

To assess the robustness of the severity categorization, we conducted a threshold sensitivity analysis by perturbing the three cutoffs by ±5 and ±10 score points while keeping the symptom score, image score, and fusion weights unchanged. In addition, we computed the distance between the final severity score and the nearest threshold. Cases falling within 5 score points of a threshold were considered borderline and flagged for cautious interpretation. This analysis was intended to evaluate whether severity labels remain stable under reasonable threshold variation and to identify cases whose classification is sensitive to the selected cutoffs.

Future work will calibrate these thresholds using real patient-level datasets containing clinician-assigned severity grades, clinical outcomes, and longitudinal disease progression. Such data-driven calibration may involve ordinal regression, receiver operating characteristic analysis, Youden-index optimization, decision-curve analysis, or agreement analysis against expert severity labels.

3.9 Reproducibility and Open Science Compliance

To ensure reproducibility and transparency, all custom code, preprocessing pipelines, model configurations, trained weights, and evaluation scripts used in this study have been publicly archived at Zenodo: https://doi.org/10.5281/zenodo.17451367

All datasets used in this study are publicly available and are cited by their original sources and licenses in the corresponding dataset sections. Image-based experiments utilize the Monkeypox Skin Images Dataset (MSID), while symptom-based classification relies on a publicly available synthetic clinical dataset.

For lesion segmentation, annotation files and derived segmentation masks were generated by the authors and are publicly available via a Kaggle repository, which is referenced in the manuscript.

3.10 Evaluation Methodology

The proposed hybrid, late-fusion AI driven framework was evaluated using a structured and modality-specific evaluation protocol to ensure robustness, generalization, and clinical relevance across all components of the system.

For symptom-based classification, model performance was assessed using a hierarchical stacking ensemble comprising six base learners. To prevent information leakage between base learners and meta-learners, a three-fold cross-validation strategy (CV = 3) was employed during the stacking process. The dataset was initially split into training and test sets using an 80/20 stratified split with a fixed random seed to ensure reproducibility. Model performance was evaluated using accuracy, confusion matrices, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC), with particular emphasis on recall for the Mpox-positive class due to its clinical importance in minimizing false negatives.

For image-based classification, seven pre-trained convolutional neural networks were fine-tuned using transfer learning and evaluated using hold-out validation. Training and validation accuracy and loss curves were monitored to assess convergence behavior and detect overfitting. Final classification decisions were obtained using a majority voting ensemble, and performance comparisons across individual models were reported using validation accuracy and loss metrics.

For lesion segmentation, a customized U-Net model was evaluated using Dice Coefficient and Intersection over Union (IoU), which are well-suited for highly imbalanced medical image segmentation tasks where lesion regions occupy a small fraction of the image. Quantitative metrics were complemented with qualitative visual inspection of predicted masks against ground truth annotations to assess spatial accuracy and boundary consistency.

Finally, severity estimation was evaluated through a deterministic scoring algorithm that integrates normalized symptom-based and image-based severity scores. Although severity scoring does not involve supervised learning, its reliability was assessed through a complete end-to-end test case demonstrating consistency between clinical symptoms, lesion burden, and final severity classification.

This multi-level evaluation strategy ensures that each component of the framework like symptom classification, image classification, lesion segmentation, and severity estimation is validated independently while also supporting reliable integration within the overall hybrid system.

3.10.1 Scope of Validation and External Generalizability

All model evaluations in this study were conducted using internal train-test or validation splits derived from the available datasets. Therefore, the reported performance values should be interpreted as internal validation results rather than evidence of external clinical generalizability. Internal validation is useful for assessing model feasibility, training stability, and potential overfitting within the same data distribution; however, it does not fully assess robustness to independent data sources, different clinical settings, different imaging devices, demographic

variation, or real-world acquisition conditions. External validation of the complete hybrid framework would require an independent patient-level multimodal dataset containing structured symptoms, confirmed mpox diagnostic labels, lesion images, lesion masks or clinician-assigned severity indicators, and clinical outcome information for the same patients. Such publicly accessible aligned datasets remain limited. For this reason, the current study should be regarded as a prototype and feasibility evaluation of the proposed late-fusion workflow.

4. Test Case: Hybrid Severity Assessment on Symptom and Image Inputs

In the preceding sections, we explained the development and workflow of the proposed hybrid framework, which integrates both symptom-based and image-based modalities for Mpox classification and severity scoring. We also demonstrated the fusion strategy used to compute the final severity score. In this section, we present a comprehensive test case in which a patient is identified as positive for mpox in both the symptom-based and image-based classifiers.

The patient scenario is as follows: The patient presents with multiple clinical symptoms suggestive of mpox and has developed visible skin lesions. It is assumed that all lesion images are captured at a standard clinical distance, ensuring that the affected skin regions and lesions are fully visible within the image frame. Following our framework, the process involves recording the symptoms, analyzing the skin lesion image, and subsequently computing the hybrid severity score.

4.1 Symptom-Based Severity Assessment

First, the patient's symptoms are collected and structured according to the WHO-recommended criteria. The recorded symptoms are summarized in **Table 4.1.1**.

Using the symptom weighting formula (Equation 3) described earlier, the symptom-based severity score was computed. Each symptom was assigned a predefined weight, and the aggregated normalized score resulted in:

Symptom-Based Severity Score: 38.83

4.2 Image-Based Severity Assessment

Next, the patient's skin lesion image was evaluated. The workflow for image-based analysis includes:

1. **Classification**:
   The input lesion image was classified as either **Mpox** or **Other** using a majority voting ensemble of seven pre-trained deep CNN models: ResNet50, MobileNetV2, EfficientNetB0, InceptionV3, Xception, VGG16, and DenseNet121.
2. **Voting Results**:
   Each model predicted the class independently, and the majority vote determined the final image classification. The classified image is shown in Figure 4.2.1, and the individual model predictions are summarized in Table 4.2.1.
3. **Lesion Segmentation:**
   Since the image was classified as positive for Mpox, lesion segmentation was performed for severity scoring. Our previously trained U-Net model was employed to predict the

lesion masks from the input image. The segmentation process yielded two outputs: A raw predicted mask as well as a segmented image with labeling, visible on Figure 4.2.2.

Connected component analysis was performed on the binarized U-Net output mask using a unified detection threshold and minimum area filtering to ensure synchronized lesion detection and counting. A total of 7 individual lesions were detected. Using the lesion count and the aggregated lesion area, the normalized image-based severity score was computed (Equation 4):

Image-Based Severity Score: 11.09

4.3 Final Severity Score and Assessment

Finally, the symptom-based and image-based scores were fused according to the weighted fusion strategy (Equation 5):

Final Severity Score: 24.96

Based on the final score, the severity level was determined by comparing against predefined thresholds (T1=20, T2=50, T3=75):

Severity Level: Moderate

The complete assessment results for this test case are summarized in Table 4.3.1.

4.3.1 Threshold Sensitivity Analysis

To examine whether the final severity category was sensitive to the selected threshold values, we repeated the severity categorization by shifting all three thresholds by ±5 and ±10 score points. The final score remained fixed at 24.96. The results are shown in Table 4.3.1.1.

The sensitivity analysis shows that the demonstrated test case is close to the mild/moderate boundary. Under the base threshold setting, the final score of 24.96 is classified as moderate; however, when T1 is increased by 5 points, the classification changes to mild. The margin from the nearest base threshold is 4.96 score points. Therefore, this case should be interpreted as a borderline mild-to-moderate case rather than a strongly moderate case. This finding highlights the importance of future clinical calibration of the threshold values.

4.4 Implementation Details

The complete implementation was carried out in Python using TensorFlow and OpenCV libraries. A pre-trained U-Net model was loaded with custom Dice coefficient and IoU metrics registered during training. Symptom severity was calculated using weighted aggregation, and lesion segmentation outputs were post-processed via connected component analysis applied consistently to both lesion detection and enumeration. A single detection threshold and a minimum lesion area constraint were employed to suppress spurious noise and ensure alignment between visual lesion marking and numerical lesion counts. The following key points were considered during computation: (a) Normalization of lesion count and lesion area before severity score computation. (b) Use of weighted fusion ($\gamma = 0.5$, $\delta = 0.5$) to balance symptom and image modalities equally. (c) Limiting final severity scores to the [0, 100] range. (d) All lesion images are assumed to be captured at a standard clinical distance, ensuring consistent lesion-to-frame proportions for reliable segmentation and severity estimation.

This test case demonstrates the operational feasibility of the proposed hybrid framework by tracing the complete pipeline from symptom input and lesion-image input to classification, segmentation, lesion quantification, score fusion, and final severity categorization. However, this test case should be interpreted as a system-level workflow demonstration rather than external clinical validation, because the available symptom and image datasets are not patient-level aligned.

4.5 Diagnostic Scenario Handling

To ensure the generalizability of the proposed framework, we defined four possible diagnostic scenarios based on symptom and image-based predictions. In typical cases where both symptom and image-based models are positive, the full severity score is used. If symptoms are present but the image-based model is negative, suggesting early or invisible lesions, half of the symptom score is considered. Conversely, if symptoms are absent but skin lesions are detected, half of the image-based score is applied to account for asymptomatic visual presentations. Finally, when both symptom and image models are negative, the severity score is set to zero, indicating a non-infected individual. This structured approach ensures the framework can adaptively handle a wide range of real-world clinical situations. The summary is given in Table 4.5.1:

4.6 System-Level Workflow Evaluation

To complement the independent evaluation of the symptom classification, image classification, and segmentation components, we performed a system-level workflow evaluation of the integrated pipeline. The purpose of this evaluation was not to claim external clinical validation, but to verify that the proposed late-fusion framework produces logically consistent outputs when all modules are executed together.

The system-level evaluation examined five aspects of the integrated workflow: (1) whether the symptom branch generated a valid symptom-based risk/severity score; (2) whether the image-classification branch generated a valid Mpox or Non-Mpox decision; (3) whether lesion segmentation was triggered only when the image-classification branch predicted Mpox; (4) whether lesion count and lesion area were successfully converted into a normalized image-based severity score; and (5) whether the symptom and image scores were fused into a bounded final score in the range [0, 100] and mapped to one of the predefined severity categories.

As detailed in the end-to-end trace of the demonstrated case (Table 4.6.2), in the positive-positive scenario, the system successfully completed all stages of the pipeline. The symptom-based severity score was 38.83. The image-classification ensemble classified the lesion image as Mpox, triggering the U-Net segmentation module. The segmentation and connected-component analysis detected 7 lesions and produced an image-based severity score of 11.09. The final fused severity score was 24.96, corresponding to a moderate category under the base threshold setting. However, because this value is close to the mild/moderate threshold, the case was further interpreted as borderline mild-to-moderate following the threshold sensitivity analysis.

We additionally evaluated the rule consistency of the framework across the four diagnostic scenarios defined in Section 4.5. Table 4.6.1 summarizes the system-level diagnostic scenario evaluation, outlining the expected pipeline behavior for each combination of symptom and image branch predictions. This scenario-level analysis confirms that the system produces bounded and interpretable severity outputs across all possible combinations of symptom- and image-based predictions. When both branches are positive, the full fused severity score is used. When only the symptom branch is positive, the framework produces a reduced symptom-only severity score. When only the image branch is positive, the framework produces a reduced image-only severity score. When both branches are negative, the final severity score is set to zero. This design prevents unsupported severity escalation when only one modality provides positive evidence. Because the current symptom and image datasets are not patient-level aligned, this system-level evaluation should be interpreted as an operational workflow and rule-consistency assessment rather than a definitive end-to-end clinical performance evaluation. A true system-level clinical evaluation will require independent patient-level multimodal data containing matched symptoms, lesion images, diagnostic confirmation, segmentation masks or clinician severity labels, and outcome information.

5. Key Findings, Limitations, and Future Research Directions

5.1 Key Findings

This study presents a prototype hybrid framework that integrates symptom-based classification, image-based lesion classification, lesion segmentation, and deterministic severity scoring for mpox assessment. The image-classification and segmentation branches demonstrated promising internal performance, while the symptom-based classifier showed moderate discrimination with high mpox-positive recall but limited Non-Mpox recall. These findings support the technical feasibility of the proposed late-fusion workflow. However, the results should be interpreted as internal feasibility evidence rather than proof of external clinical validity. A complete end-to-end test case demonstrated the operational workflow of the system, but external validation using independent patient-level multimodal data is required before clinical deployment.

5.2 Limitations

Despite encouraging internal results, this study has several important limitations. First, all reported experiments were conducted using internal train-test or validation splits of the available datasets. Although internal validation is useful for assessing feasibility and identifying overfitting within the same data distribution, it does not establish external generalizability across independent clinical settings, imaging devices, demographic groups, or disease presentations. Therefore, the reported results should be interpreted as internal feasibility evidence rather than proof of real-world clinical robustness.

Second, the complete hybrid framework has not yet been externally validated using patient-level multimodal data. Such validation would require independent datasets linking structured symptoms, confirmed diagnostic labels, lesion images, segmentation masks or clinician-assigned severity grades, and clinical outcomes from the same individuals. Because the symptom and image datasets used in this study are not patient-level aligned, the proposed framework performs

decision-level late fusion rather than learning direct inter-modality relationships from matched patient records.

Third, the symptom-based classifier achieved only moderate discriminative performance, with an AUC of approximately 0.70–0.71. Although the model showed high recall for mpox-positive cases, recall for Non-Mpox cases was low, indicating limited specificity and a tendency toward false-positive predictions. This may reduce reliability when differentiating mpox from clinically similar Non-Mpox conditions. Therefore, the symptom classifier should be interpreted only as a preliminary risk-stratification component and not as a standalone diagnostic model.

Fourth, the symptom-based classifier was trained and evaluated using a synthetic structured dataset. Synthetic data cannot fully represent real-world variability in symptom co-occurrence, disease stage, immune status, vaccination history, reporting noise, comorbidities, or overlap with other dermatological and sexually transmitted infections. External validation using real patient-level symptom records is required before clinical interpretation.

Fifth, the segmentation dataset was limited to 89 original manually annotated images expanded to 445 samples through augmentation. Although augmentation improves robustness to selected transformations, it cannot replace genuine clinical diversity across patients, skin tones, anatomical sites, lesion morphologies, imaging devices, or disease stages. External validation using larger expert-annotated segmentation datasets is required before clinical deployment.

Sixth, the severity scoring system is deterministic and guideline-informed, but it has not yet been prospectively validated against clinician-assigned severity grades or patient outcomes. In addition, image-based severity estimation assumes relatively standardized image acquisition conditions. Images captured at varying distances, under poor illumination, or with partial lesion visibility may affect lesion segmentation, lesion counting, and final severity estimation.

The severity thresholds used in this study are preliminary framework-specific cutoffs applied to a normalized 0–100 composite score. Although they provide an interpretable method for mapping the final score into mild, moderate, severe, and serious categories, they have not yet been optimized using clinician-assigned severity grades or patient outcomes. The threshold sensitivity analysis showed that cases close to a cutoff may change category under modest threshold perturbation. Therefore, severity categories near threshold boundaries should be interpreted cautiously and considered borderline until the thresholds are empirically calibrated on real-world patient-level data.

The system-level evaluation presented in this study demonstrates operational feasibility and rule consistency of the integrated pipeline, but it does not establish clinical utility. Because the symptom and image datasets are not patient-level aligned, the current study cannot calculate true end-to-end diagnostic accuracy or severity-classification accuracy for the full multimodal system. Future validation will require independent multimodal clinical datasets containing matched symptoms, lesion images, laboratory confirmation, clinician-assigned severity grades, and outcome information.

5.3 Future Research Directions

Future work should focus on expanding datasets with greater ethnic and temporal diversity, incorporating longitudinal data for dynamic severity tracking, integrating explainable AI for improved transparency, piloting clinical deployment through mobile or telemedicine platforms, and adapting the modular framework to other dermatological infectious diseases for broader epidemic readiness. Future studies will also aim to incorporate scale-invariant normalization techniques to mitigate the effects of variable image acquisition conditions.

Future work will focus on validating the proposed framework using real-world, patient-level multimodal datasets containing structured symptoms, confirmed diagnostic labels, lesion images, clinician-assigned severity grades, and outcome information. Larger externally sourced segmentation datasets with expert-verified masks should be used to assess generalization across skin tones, lesion stages, anatomical regions, and imaging conditions. In addition, prospective clinical evaluation will be necessary to calibrate the severity scoring thresholds and determine whether the framework can reliably support triage and decision-making in real healthcare settings.

Future studies should perform external validation using independent datasets collected from different institutions, geographic regions, imaging conditions, and patient populations. In particular, validation should include patient-level multimodal datasets that link structured symptoms, confirmed diagnostic labels, lesion images, clinician-assigned severity grades, and clinical outcomes. For the symptom-based classifier, future work should focus on improving Non-Mpox recall and specificity through threshold calibration, richer clinical features, inclusion of differential diagnoses, and validation on real-world patient records.

Future work should empirically calibrate the severity thresholds using patient-level multimodal datasets with clinician-assigned severity labels and clinical outcome information. Threshold optimization may be performed using ordinal regression, ROC-based cut-point analysis, decision-curve analysis, and agreement testing against expert severity ratings. In addition, future studies should perform true end-to-end system-level validation on independent clinical datasets in which symptoms, lesion images, segmentation masks or clinician severity grades, and diagnostic labels are available for the same patients.

6. Conclusion

In this study, we proposed a comprehensive hybrid framework for Mpox severity assessment by integrating symptom-based and image-based modalities. The symptom-based component utilizes a carefully weighted scoring system derived from clinical guidelines to quantify patient-reported symptoms, while the image-based component employs a reliable deep learning pipeline combining ensemble classification and U-Net segmentation for lesion detection and quantification. By leveraging seven pre-trained convolutional neural networks (CNNs) through majority voting, we ensured reliable classification of skin lesions as Mpox or Non-Mpox, followed by precise lesion segmentation for severity scoring. A key strength of our framework lies in its dual-modality approach, allowing for complementary insights from both clinical symptoms and visual lesion characteristics. The final severity score is computed by fusing these two modalities, ensuring a holistic and patient-specific evaluation. We validated our system

through a detailed test case, demonstrating the end-to-end workflow from patient data input to final severity determination, categorized into clinically relevant levels: mild, moderate, severe, and serious. The hybrid design provides a quantifiable severity index, which can assist clinicians in prioritizing care, especially in outbreak scenarios where rapid triage is critical. The integration of segmentation-based lesion quantification further strengthens the model's utility by offering interpretable and explainable visual outputs alongside numerical scores. Future work could extend this approach by incorporating temporal progression analysis of symptoms and lesions, adding multimodal patient history data, or fine-tuning the model on larger, more diverse datasets to improve generalizability across different populations and regions. Nonetheless, the current framework offers a significant step toward automated, scalable, and interpretable Mpox severity assessment, addressing a critical gap in digital health preparedness for emerging infectious diseases.

References

[1] Moore MJ, Rathish B, Zahra F. Mpox (Monkeypox) [Updated 2023 May 3]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK574519/

[2] Alakunle, E., Moens, U., Nchinda, G., & Okeke, M. I. (2020). Monkeypox Virus in Nigeria: Infection Biology, Epidemiology, and Evolution. *Viruses*, *12*(11), 1257. https://doi.org/10.3390/v12111257

[3] Kugelman, J. R., Johnston, S. C., Mulembakani, P. M., Kisalu, N., Lee, M. S., Koroleva, G., McCarthy, S. E., Gestole, M. C., Wolfe, N. D., Fair, J. N., Schneider, B. S., Wright, L. L., Huggins, J., Whitehouse, C. A., Wemakoy, E. O., Muyembe-Tamfum, J. J., Hensley, L. E., Palacios, G. F., & Rimoin, A. W. (2014). Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo. *Emerging infectious diseases*, *20*(2), 232–239. https://doi.org/10.3201/eid2002.130118

[4] Sklenovská, N., & Van Ranst, M. (2018). Emergence of Monkeypox as the Most Important Orthopoxvirus Infection in Humans. *Frontiers in public health*, *6*, 241. https://doi.org/10.3389/fpubh.2018.00241

[5] Monkeypox. (2011). *Releve epidemiologique hebdomadaire*, *86*(41), 448–451.

[6] Hutson, C. L., Carroll, D. S., Gallardo-Romero, N., Drew, C., Zaki, S. R., Nagy, T., Hughes, C., Olson, V. A., Sanders, J., Patel, N., Smith, S. K., Keckler, M. S., Karem, K., & Damon, I. K. (2015). Comparison of Monkeypox Virus Clade Kinetics and Pathology within the Prairie Dog

Animal Model Using a Serial Sacrifice Study Design. *BioMed research international*, *2015*, 965710. https://doi.org/10.1155/2015/965710
[7] McCollum, A. M., & Damon, I. K. (2014). Human monkeypox. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, *58*(2), 260–267. https://doi.org/10.1093/cid/cit703
[8] Weaver, J. R., & Isaacs, S. N. (2008). Monkeypox virus and insights into its immunomodulatory proteins. *Immunological reviews*, *225*, 96–113. https://doi.org/10.1111/j.1600-065X.2008.00691.x
[9] Chauhan, R. P., Fogel, R., & Limson, J. (2023). Overview of Diagnostic Methods, Disease Prevalence and Transmission of Mpox (Formerly Monkeypox) in Humans and Animal Reservoirs. *Microorganisms*, *11*(5), 1186. https://doi.org/10.3390/microorganisms11051186
[10] Jain, N., Umar, T. P., Sayad, R., Mokresh, M. E., Tandarto, K., Siburian, R., Liana, P., Laivacuma, S., & Reinis, A. (2024). Monkeypox Diagnosis in Clinical Settings: A Comprehensive Review of Best Laboratory Practices. *Advances in experimental medicine and biology*, *1451*, 253–271. https://doi.org/10.1007/978-3-031-57165-7_16
[11] K. D. Akin, C. Gurkan, A. Budak, and H. Karatas, "Classification of Monkeypox Skin Lesion using the Explainable Artificial Intelligence Assisted Convolutional Neural Networks," European Journal of Science and Technology, no. 40, pp. 106–110, Sep. 2022, doi: 10.31590/ejosat.1171816.
[12] M. M. Ahsan, T. A. Abdullah, M. S. Ali, F. Jahora, M. K. Islam, A. G. Alhashim, and K. D. Gupta, "Transfer Learning and Local Interpretable Model-Agnostic Based Visual Approach in Monkeypox Disease Detection and Classification: A Deep Learning Insights," arXiv preprint arXiv:2211.05633, 2022.
[13] Khan, Saddam Hussain, and Rashid Iqbal. "RS-FME-SwinT: A Novel Feature Map Enhancement Framework Integrating Customized SwinT with Residual and Spatial CNN for Monkeypox Diagnosis." *arXiv preprint arXiv:2410.01216* (2024).
[14] Lakshmi, M.; Das, R. Classification of Monkeypox Images Using LIME-Enabled Investigation of Deep Convolutional Neural Network. Diagnostics 2023, 13, 1639. https://doi.org/10.3390/diagnostics13091639
[15] Jaradat AS, Al Mamlook RE, Almakayeel N, Alharbe N, Almuflih AS, Nasayreh A, Gharaibeh H, Gharaibeh M, Gharaibeh A, Bzizi H. Automated Monkeypox Skin Lesion Detection Using Deep Learning and Transfer Learning Techniques. Int J Environ Res Public Health. 2023 Mar 1;20(5):4422. doi: 10.3390/ijerph20054422. PMID: 36901430; PMCID: PMC10001976.
[16] Karagoz A, Tombuloglu H, Alsaeed M, Tombuloglu G, AlRubaish AA, Mahmoud A, Smajlović S, Ćordić S, Rabaan AA, Alsuhaimi E. Monkeypox (mpox) virus: Classification, origin, transmission, genome organization, antiviral drugs, and molecular diagnosis. J Infect Public Health. 2023 Apr;16(4):531-541. doi: 10.1016/j.jiph.2023.02.003. Epub 2023 Feb 9. PMID: 36801633; PMCID: PMC9908738.

[17] Sahin, V.H., Oztel, I. & Yolcu Oztel, G. Human Monkeypox Classification from Skin Lesion Images with Deep Pre-trained Network using Mobile Application. J Med Syst 46, 79 (2022). https://doi.org/10.1007/s10916-022-01863-7
[18] Demir, F.B., Baygin, M., Tuncer, I. et al. MNPDenseNet: Automated Monkeypox Detection Using Multiple Nested Patch Division and Pretrained DenseNet201. Multimed Tools Appl 83, 75061–75083 (2024). https://doi.org/10.1007/s11042-024-18416-4
[19] Abdellatef, E., Ismail, A.H., Fath Allah, M.I. et al. Leveraging convolutional neural networks and hashing techniques for the secure classification of monkeypox disease. Sci Rep 14, 26579 (2024). https://doi.org/10.1038/s41598-024-75030-y
[20] Altun, M.; Gürüler, H.; Özkaraca, O.; Khan, F.; Khan, J.; Lee, Y. Monkeypox Detection Using CNN with Transfer Learning. Sensors 2023, 23, 1783. https://doi.org/10.3390/s23041783
[21] Titanji BK, Tegomoh B, Nematollahi S, Konomos M, Kulkarni PA. Monkeypox: A Contemporary Review for Healthcare Professionals. Open Forum Infect Dis. 2022 Jun 23;9(7):ofac310. doi: 10.1093/ofid/ofac310. PMID: 35891689; PMCID: PMC9307103.
[22] Huhn GD, Bauer AM, Yorita K, Graham MB, Sejvar J, Likos A, Damon IK, Reynolds MG, Kuehnert MJ. Clinical characteristics of human monkeypox, and risk factors for severe disease. Clin Infect Dis. 2005 Dec 15;41(12):1742-51. doi: 10.1086/498115. Epub 2005 Nov 11. PMID: 16288398.
[23] Reynolds MG, Yorita KL, Kuehnert MJ, Davidson WB, Huhn GD, Holman RC, Damon IK. Clinical manifestations of human monkeypox influenced by route of infection. J Infect Dis. 2006 Sep 15;194(6):773-80. doi: 10.1086/505880. Epub 2006 Aug 8. PMID: 16941343.
[24] Hatmal MM, Al-Hatamleh MAI, Olaimat AN, Ahmad S, Hasan H, Ahmad Suhaimi NA, Albakri KA, Abedalbaset Alzyoud A, Kadir R, Mohamud R. Comprehensive literature review of monkeypox. Emerg Microbes Infect. 2022 Dec;11(1):2600-2631. doi: 10.1080/22221751.2022.2132882. PMID: 36263798; PMCID: PMC9627636.
[25] Mustafa Altindis, Edmond Puca, Laidon Shapo, Diagnosis of monkeypox virus – An overview, Travel Medicine and Infectious Disease,Volume 50,2022, 102459, ISSN 1477-8939, https://doi.org/10.1016/j.tmaid.2022.102459.
[26] Jezek Z, Szczeniowski M, Paluku KM, Mutombo M. Human monkeypox: clinical features of 282 patients. J Infect Dis. 1987 Aug;156(2):293-8. doi: 10.1093/infdis/156.2.293. PMID: 3036967.
[27] Harapan H, Ophinni Y, Megawati D, Frediansyah A, Mamada SS, Salampe M, Bin Emran T, Winardi W, Fathima R, Sirinam S, Sittikul P, Stoian AM, Nainu F, Sallam M. Monkeypox: A Comprehensive Review. Viruses. 2022 Sep 29;14(10):2155. doi: 10.3390/v14102155. PMID: 36298710; PMCID: PMC9612348.
[28] Thornhill JP, Barkati S, Walmsley S, Rockstroh J, Antinori A, Harrison LB, Palich R, Nori A, Reeves I, Habibi MS, Apea V, Boesecke C, Vandekerckhove L, Yakubovsky M, Sendagorta E, Blanco JL, Florence E, Moschese D, Maltez FM, Goorhuis A, Pourcher V, Migaud P, Noe S, Pintado C, Maggi F, Hansen AE, Hoffmann C, Lezama JI, Mussini C, Cattelan A, Makofane K, Tan D, Nozza S, Nemeth J, Klein MB, Orkin CM; SHARE-net Clinical Group. Monkeypox

Virus Infection in Humans across 16 Countries - April-June 2022. N Engl J Med. 2022 Aug 25;387(8):679-691. doi: 10.1056/NEJMoa2207323. Epub 2022 Jul 21. PMID: 35866746.
[29] Rizk JG, Lippi G, Henry BM, Forthal DN, Rizk Y. Prevention and Treatment of Monkeypox. Drugs. 2022 Jun;82(9):957-963. doi: 10.1007/s40265-022-01742-y. Epub 2022 Jun 28. Erratum in: Drugs. 2022 Aug;82(12):1343. doi: 10.1007/s40265-022-01767-3. PMID: 35763248; PMCID: PMC9244487
[30] Arshed, M.A.; Rehman, H.A.; Ahmed, S.; Dewi, C.; Christanto, H.J. A 16 × 16 Patch-Based Deep Learning Model for the Early Prognosis of Monkeypox from Skin Color Images. Computation 2024, 12, 33. https://doi.org/10.3390/computation12020033
[31] Almufareh, M.F.; Tehsin, S.; Humayun, M.; Kausar, S. A Transfer Learning Approach for Clinical Detection Support of Monkeypox Skin Lesions. Diagnostics 2023, 13, 1503. https://doi.org/10.3390/diagnostics13081503
[32] Uysal, F. Detection of Monkeypox Disease from Human Skin Images with a Hybrid Deep Learning Model. Diagnostics 2023, 13, 1772. https://doi.org/10.3390/diagnostics13101772
[33] Yolcu Oztel, G. Vision transformer and CNN-based skin lesion analysis: classification of monkeypox. Multimed Tools Appl 83, 71909–71923 (2024). https://doi.org/10.1007/s11042-024-19757-wMonkeypox. Computers in Biology and Medicine, 152, 106332.
[34] Dahiya, N., Sharma, Y.K., Rani, U. et al. Hyper-parameter tuned deep learning approach for effective human monkeypox disease detection. Sci Rep 13, 15930 (2023). https://doi.org/10.1038/s41598-023-43236-1
[35] Bala, Diponkor; Hossain, Md Shamim (2023), "Monkeypox Skin Images Dataset (MSID)", Mendeley Data, V6, doi: 10.17632/r9bfpnvyxr.6
[36]Muhamad Ahmed. (2022). Monkey-Pox PATIENTS Dataset. [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/4271503
[37] https://www.kaggle.com/datasets/jewelmd/mpox-classification-augmneted-images
[38] https://www.kaggle.com/datasets/jewelmd/mpox-lesion-segmentation-suitable-for-unet
[39] McNeil AJ, House DW, Mbala-Kingebeni P, Mbaya OT, Dodd LE, Cowen EW, Nussenblatt V, Bonnett T, Chen Z, Saknite I, Dawant BM, Tkaczyk ER. Counting Monkeypox Lesions in Patient Photographs: Limits of Agreement of Manual Counts and Artificial Intelligence. J Invest Dermatol. 2023 Feb;143(2):347-350.e4. doi: 10.1016/j.jid.2022.08.044. Epub 2022 Sep 16. PMID: 36116509; PMCID: PMC9534148.
[40] World Health Organization. (2024, August 26). *Mpox (monkeypox)*. World Health Organization. Retrieved April 26, 2025, from https://www.who.int/news-room/fact-sheets/detail/mpox
[41] World Health Organization. (2022, June 10). Clinical management and infection prevention and control for monkeypox: Interim rapid response guidance, 10 June 2022 (WHO/MPX/Clinical and IPC/2022.1). World Health Organization. https://www.who.int/publications/i/item/WHO-MPX-Clinical-and-IPC-2022.1
[42] Abdullah Reda, Lina Hemmeda, Aml M. Brakat, Ranjit Sah, Amr Ehab El-Qushayri, The clinical manifestations and severity of the 2022 monkeypox outbreak among 4080 patients,

Travel Medicine and Infectious Disease, Volume 50, 2022, 102456, ISSN 1477-8939, https://doi.org/10.1016/j.tmaid.2022.102456. (https://www.sciencedirect.com/science/article/pii/S1477893922002022)
[43] Das, A., et al. (2026). Monkeypox Detection and Classification Using Deep Learning with Attention and Ensemble Modeling. *IEEE Xplore*. https://ieeexplore.ieee.org/document/11366870/
[44] Askale, G.T., et al. (2026). A customized MobileNetV2-based lightweight CNN for monkeypox detection and classification. *Scientific Reports*. https://doi.org/10.1038/s41598-026-35871-1
[45] Vandana, V., et al. (2026). MpoxSegNet for multiclass monkeypox segmentation and classification. *SN Computer Science*. https://doi.org/10.1007/s44163-026-00884-2
[46] Zhang, Z., et al. (2026). Attention-enhanced hybrid deep learning framework for Monkeypox skin lesion classification. *Electronic Research Archive*, 34(2), 738-776.
[47] Sharma, C., et al. (2026). A Lightweight Attention CNN Model for Monkeypox Detection with Evolutionary Optimization. *Journal of Transformative Technologies and Sustainable Development*.
[48] Alghoraibi, S., et al. (2025–2026). ITMA'INN: Vision Transformers and MobileViT for Mpox detection and real-time monitoring. *Journal of King Saud University - Computer and Information Sciences*.
[49] Onyema, E.M., et al. (2025–2026). Deep learning model for hair artifact removal and Mpox skin lesion analysis and detection. *Scientific Reports*. https://doi.org/10.1038/s41598-025-05324-2
[50] Ogoina, D., et al. (2024). The 2024 Mpox Clade I outbreak: Clinical characteristics and outcomes in endemic regions. *The Lancet Infectious Diseases*, 24(5), 512-521.
[51] Mitjà, O., et al. (2024). Clinical presentation and progression of Mpox in diverse populations: A global cohort analysis. *New England Journal of Medicine*, 390(14), 1281-1292.
[52] Orkin, C., et al. (2024). Efficacy of Tecovirimat for Mpox in the PLATINUM-S randomized controlled trial. *JAMA*, 332(5), 389-398.
[53] World Health Organization. (2024). Mpox (Monkeypox) - Public Health Emergency of International Concern (PHEIC) declaration regarding Clade Ib. *WHO Newsroom*.
[54] WHO Strategic Advisory Group of Experts (SAGE). (2025). Updated clinical management and severity guidelines for Mpox variants. *WHO Guidelines*.
[55] Kamara, A., et al. (2025). Machine learning-driven forecasting of Mpox outbreaks using epidemiological and mobility data. *Epidemiology and Infection*, 153, e45.
[56] Smith, J. R., et al. (2024). Real-time digital surveillance and AI-driven vaccine efficacy monitoring for Mpox. *Vaccine*, 42(15), 2340-2348.
[57] Garcia, M., et al. (2024). Utilizing large language models for automated Mpox symptom triage and patient routing. *Journal of the American Medical Informatics Association*, 31(7), 1450-1458.

[58] Bianchi, V., et al. (2024). A multimodal AI framework integrating symptoms and imaging for dermatological emergencies. *Nature Communications*, 15, 4321.
[59] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).
[60] Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
[61] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).
[62] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).
[63] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
[64] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
[65] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).
[66] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015* (pp. 234-241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28